

# A coherence enhancing penalty for Diffusion MRI: regularizing property and discrete approximation

T. Hohage\*      C. Rügge\*

December 4, 2014

Processing of Diffusion MRI data obtained from High Angular Resolution measurements consists of a series of steps, starting with the estimation of an orientation distribution function (ODF), which is then used as input for e.g. tractography algorithms. It is important that ODF reconstruction methods yield accurate, coherent ODFs, in particular for low SNR or coarsely sampled data sets. As the diffusion process is modelled independently in each voxel, reconstructions are often carried out for each voxel separately, disregarding the observation that neighboring voxels are often quite similar if they belong to the same fiber structure. There are surprisingly few approaches that make use of this kind of spatial regularity to improve coherence and stability of the reconstruction. In this work, we focus on a variation of a method proposed by Reisert and Kiselev based on the concept of fiber continuity. The method has already been shown to yield good numerical results, but has not yet been analyzed theoretically. Under suitable smoothness assumptions, we apply results on constrained Tikhonov-type regularization with approximate operator to show convergence of reconstructions from discrete, noisy data for linear forward models. Further, we numerically illustrate the performance of the method on phantom and *in-vivo* data.

## 1 Introduction

Diffusion weighted MRI (DW-MRI) is a non-invasive method to measure the movements of water molecules in biological tissue. Using the fact that water diffusion is mainly directed along nerve fibers, not perpendicular to them, it can be used to resolve the fibrous structure of brain white matter, with a wide range of applications in both medicine and neuro-science. DW measurements acquire a number of full 3-dimensional MRI volumes with varying diffusion sensitizing gradients encoding the diffusion in different directions. From this, one tries to infer information on the diffusive properties of the tissue as parametrized by a suitable physical model of the diffusion process. The most widely-used model in this regard, the tensor model,

---

\*Institute for Numerical and Applied Mathematics, University of Göttingen, Germany

is based on the assumption of anisotropic Gaussian diffusion parametrized by a diffusion tensor in each voxel. While being quite successful, its main drawback is the inherent inability to resolve more than one diffusion direction per voxel. However, depending on the resolution, as much as a third of all voxels can contain multiple fibers.

More refined approaches often replace the diffusion tensor by an *orientation distribution function* (ODF) measuring either — depending on the underlying model — the diffusion probability or the density of fibers per direction for each voxel of the volume. This requires a larger number of diffusion sensitizing gradients than for the tensor model, often termed as High Angular Resolution Diffusion Imaging (HARDI).

Most models describe diffusion in each voxel separately, and for efficiency, reconstructions are usually simply carried out voxel-wise. This approach however neglects spatial coherence, i.e. the fact that ODFs in nearby voxels are often similar if they belong to the same fiber bundle. Incorporating spatial coherence into the reconstruction algorithm can potentially improve the results significantly. Spatial regularization was first investigated in [6] where similarity is measured by comparing the entire ODFs in nearby voxels. Smoothing methods that take into account the underlying structure of the domain  $\Omega \times S^2$  have been suggested for example in [2], where linear and non-linear diffusion filters are applied to the reconstructed ODF in a post-processing step, and in [1], where adaptive smoothing is performed on the HARDI data prior to reconstruction.

In this paper we will study the concept of *fiber continuity* (FC) as presented in [15], which uses related ideas to regularize the ODF reconstruction. Here similarity between ODFs is *local* also in the orientational part and only compares voxels along fibers instead of isotropically. This way, smoothness information can extend for example from single-fiber voxels into adjacent crossings despite the sudden appearance of perpendicular structures which violate the *global* similarity of the respective ODFs. The approach is based on the assumption that curvature of the fiber bundles is not too large, and that the point  $(x, u) \in \Omega \times S^2$  belongs to a fibrous structure through  $x$ , directed along  $u$ . So for sufficiently small step-lengths  $\tau > 0$ , the ODF field  $\psi: \Omega \times S^2 \rightarrow \mathbb{R}$  assigning to each voxel  $x$  in the spatial domain  $\Omega$  the ODF  $\psi(x, \cdot)$  fulfills

$$\psi(x + \tau u, u) \simeq \psi(x, u).$$

This assumption is employed in a Tikhonov-type regularization scheme, with the anisotropic regularization term

$$\int_{\Omega} \int_{S^2} |u^T \text{grad}_x \psi(x, u)|^2 du dx. \quad (1)$$

Although this regularization scheme leads to good results, its theoretical properties have not been studied yet. In particular, compactness properties, which are needed to establish convergence of discrete approximations, are not obvious since the forward operator is not smoothing in the spatial variable  $x$ . The aim of this work is to analyze the convergence of a spatial regularization method similar to the one above for noisy, discrete data. This will be achieved by studying the properties of a non-standard Sobolev-type space constructed from the anisotropic derivative.

The paper is organized as follows: in §2, we describe the diffusion model and the regularization method, and we formally introduce the Sobolev-type space  $H_{\text{fb}}$  used in this method. In §3, we review some convergence results for Tikhonov regularization with approximate operators, with focus on approximation by finite dimensional operators, i.e. discretization. In §4, the space  $H_{\text{fb}}$  will be further investigated, and we show the main result of this paper, a

compact embedding theorem that allows us to apply the convergence results from §3 to our method. Finally, §5 and 6 describe some implementation details and show some numerical results, comparing the performance of the method to standard unregularized reconstructions.

## 2 Reconstruction scheme

Modelling of the DW signal is a difficult problem in itself since the diffusive properties of white matter can be quite complex. We will use the spherical deconvolution (SD) model that was introduced in [17], which is particularly attractive due to its conceptual simplicity and linearity. It has been successfully employed in both phantom and *in-vivo* studies.

The model describes the data as  $S^2$ -convolution of the (fiber) ODF  $\psi: \Omega \times S^2 \rightarrow \mathbb{R}$  with a single-fiber response function  $k: [-1, 1] \rightarrow \mathbb{R}$  that is assumed to be identical throughout the volume:

$$T\psi(x, q) := \int_{S^2} k(q^T u) \psi(x, u) du, \quad q \in S^2.$$

The response function  $k$  can either be estimated from the data in a pre-processing step or fixed a priori by modelling it e.g. as a Gaussian.

A problem of this model, apart from suffering from instability due to ill-posedness, is that it can lead to ODFs with negative values, which makes interpretation as a diffusion probability or fiber density impossible. Therefore, one often includes non-negativity constraints  $\psi \geq 0$ . In the context of SD, non-negativity constraints have been introduced in [18], the resulting method being called constrained SD (CSD). Adding the fiber continuity based spatial regularization strategy introduced in [15], the method can then be written as

$$\operatorname{argmin}_{\psi \geq 0} \left( \|T\psi - S^\delta\|^2 + \gamma \int_{\Omega \times S^2} |u^T \operatorname{grad} \psi(x, u)|^2 dx du \right), \quad (2)$$

where  $\gamma > 0$  is a regularization parameter and  $S^\delta$  is the given noisy data with

$$\|S - S^\delta\| \leq \delta$$

for  $\delta > 0$  and exact data  $S = T\psi^\dagger$ . Note that in the context of discretization,  $\delta$  quantifies both measurement errors and errors caused by interpolating the data from a finite sampling grid to all of  $\Omega \times S^2$ . Our approach differs from this slightly: instead of (2), we solve

$$\begin{aligned} \operatorname{argmin}_{\psi \geq 0} \left( \|T\psi - S^\delta\|^2 + \alpha \|\psi\|^2 - \beta \langle \psi, \Delta_{S^2} \psi \rangle \right. \\ \left. + \gamma \int_{\Omega \times S^2} |u^T \operatorname{grad} \psi(x, u)|^2 dx du \right). \end{aligned} \quad (3)$$

This is both due to theoretical reasons to be detailed below, and the numerical experience that the additional angular smoothness introduced by the  $\Delta_{S^2}$ -term improves stability of the reconstruction and reduces artifacts for curved structures. The occurrence of these artifacts is not surprising since the regularization term (1) was derived for locally straight fibers. However, angular smoothness also tends to blur structures and limit the achievable angular resolution of the ODFs. Therefore, the parameter  $\beta$  has to be chosen carefully.

For a rigorous treatment of this optimization problem, we first need to define the function space in which the problem is posed. The spatial domain  $\Omega$  will be simply taken as a cube in  $\mathbb{R}^3$ ,

$$\Omega = \prod_{i=1}^3 \left( -\frac{L_i}{2}, \frac{L_i}{2} \right),$$

with  $L \in \mathbb{R}_+^3$ . In  $L^2(\Omega \times S^2)$ , introduce the orthonormal basis functions

$$\eta_{klm} := \xi_k \otimes Y_{lm}, \quad (4)$$

where  $\xi_k: \Omega \rightarrow \mathbb{C}$  for  $k \in \mathbb{Z}^3$  is the trigonometric orthonormal basis on  $\Omega$  and  $Y_{lm}: S^2 \rightarrow \mathbb{C}$  for  $l \in \mathbb{N}_0$  and  $m = -l, \dots, l$  are the Spherical Harmonics on  $S^2$ . Let  $C_{\text{per}}^\infty(\Omega \times S^2)$  be the space of infinitely differentiable functions on  $\mathbb{R}^3 \times S^2$  the spatial part of which is periodic with period  $L$ . For  $\psi \in C_{\text{per}}^\infty(\Omega \times S^2)$ , we denote by  $\text{grad } \psi$  and  $\text{Grad } \psi$  the derivatives with respect to the spatial and orientational parts, respectively. Moreover, we extend this notation to functions  $\psi \in L^2(\Omega \times S^2)$  using the basis  $\eta_{klm}$ , i.e.

$$\text{grad } \psi = \sum_{klm} \langle \psi, \eta_{klm} \rangle (\text{grad } \xi_k) \otimes Y_{lm}$$

and

$$\text{Grad } \psi = \sum_{klm} \langle \psi, \eta_{klm} \rangle \xi_k \otimes (\text{Grad } Y_{lm}),$$

if the series converge in  $L^2(\Omega \times S^2)^3$ . Here, we view  $\text{Grad } Y_{lm}$  as a map  $S^2 \rightarrow \mathbb{C}^3$  with  $u^T \text{Grad } Y_{lm}(u) = 0$  for all  $u \in S^2$ . For notational simplicity, the function  $\Omega \times S^2 \rightarrow S^2$ ,  $(x, u) \mapsto u$  will be simply denoted by  $u$  in the following.

**Definition 2.1.** The **fiber derivative** of  $\psi \in L^2(\Omega \times S^2)$  is defined as

$$D_{\text{fib}} \psi := \sum_{klm} \langle \psi, \eta_{klm} \rangle \sum_{i=1}^3 \frac{\partial \xi_k}{\partial x_i} \otimes (u_i Y_{lm}),$$

if the series converges in  $L^2(\Omega \times S^2)$ . The **fiber space**  $H_{\text{fib}}(\Omega \times S^2)$  is defined as the space of all  $\psi \in L^2(\Omega \times S^2)$  for which

$$\|\psi\|_{\text{fib}} := (\|\psi\|^2 + \|D_{\text{fib}} \psi\|^2 + \|\text{Grad } \psi\|^2)^{\frac{1}{2}} \quad (5)$$

is finite.

If  $\text{grad } \psi$  exists, then  $D_{\text{fib}} \psi = u^T \text{grad } \psi$  is just the operator in (2). Note that we could have introduced scaling factors into the definition of  $\|\cdot\|_{\text{fib}}$  as in (3), but for the theoretical analysis it is sufficient to omit them. The convolution operator  $T$  is now viewed as a map  $T: H_{\text{fib}}(\Omega \times S^2) \rightarrow L^2(\Omega \times S^2)$ . The properties of this space will be studied in §4.

### 3 Regularization with discretization

In this section, we briefly review some results on regularization with discretization and prove a variant of a result in [13]. Discretization of (3) is performed as a projection method by

introducing orthogonal projections  $P_h$  in  $H_{\text{fib}}(\Omega \times S^2)$  and  $Q_h$  in  $L^2(\Omega \times S^2)$ , with some discretization parameter  $h > 0$ , and replacing  $T$  with

$$T_h := Q_h T P_h,$$

i.e.

$$\psi_{\alpha,h}^\delta = \operatorname{argmin} \{ \|T_h \psi - S^\delta\|^2 + \alpha \|\psi\|_{\text{fib}}^2 : \psi \in H_{\text{fib}}(\Omega \times S^2), \psi \geq 0 \}. \quad (6)$$

There is a number of papers on regularization with discretization in the unconstrained case, we only cite [14]. Under a Hölder-type source condition,

$$\psi^\dagger \in \mathcal{R}((T^*T)^\mu) \quad (7)$$

and with a suitable parameter choice, one obtains the convergence rate

$$\|\psi_{\alpha,h}^\delta - \psi^\dagger\| = \mathcal{O}(\delta^{\frac{2\mu}{2\mu+1}} + \|T - T_h\|^{2\mu}) \quad (8)$$

for  $0 < \mu \leq \frac{1}{2}$ . Unfortunately, the proof relies on spectral theory and therefore does not generalize to the constrained case. Moreover, the convergence  $T_h \rightarrow T$  requires  $T$  to be compact, which is not immediately clear in our case: at least viewed as an operator from  $L^2(\Omega \times S^2)$  to itself, the operator  $T$  is not compact due to the spatial part.

To address the first problem, spectral theory can be replaced by variational techniques in regularization theory developed in the last decade. An important part in this approach consists of replacing the spectral source condition (7) by a condition in the form of a variational inequality, a so-called variational source condition. We will assume that

$$\beta \|\psi - \psi^\dagger\|^2 \leq \|\psi\|^2 - \|\psi^\dagger\|^2 + \phi(\|T(\psi - \psi^\dagger)\|) \quad \text{for all } \psi \in C, \quad (9)$$

some  $\beta > 0$ , and a strictly increasing, concave function  $\phi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\phi(0) = 0$ . Here,  $C$  is the convex constraint set.

It has been shown in [10] that the spectral Hölder condition (7) implies the variational source condition (9) with  $\phi(t) = t^{\frac{2\mu}{2\mu+1}}$  for any set  $C$ , i.e. even if  $C$  is the whole space. For closed, convex sets  $C$  and  $\phi(t) \sim \sqrt{t}$ , (9) is equivalent (c.f. [5]) to the projected source condition

$$\psi^\dagger \in \mathcal{R}(P_C T^*)$$

that is well-known (c.f. [3]) for this type of problem.

In [13], convergence was shown for non-linear Tikhonov regularization with approximate operators employing a variational smoothness assumption. The following result is a small modification — with almost identical proof — of [13, Theorem 3.1], specializing it to the case of linear operators but extending it to unbounded constraint sets  $C$ , for which the assumption  $\sup_{\psi \in C} \|T(\psi) - T_h(\psi)\| \rightarrow 0$  in [13] is typically not satisfied.

**Theorem 3.1.** *Let  $X, Y$  be Hilbert spaces,  $C \subset X$  closed and convex,  $T, T_h: X \rightarrow Y$  linear operators with  $\|T - T_h\| \leq \eta_h$ , and assume that the variational smoothness assumption (9) hold true. Given  $S^\delta \in Y$  with  $\|S^\delta - T\psi^\dagger\| \leq \delta$ , let  $\psi_{\alpha,h}^\delta \in C$  be the (unique) minimizer of*

$$C \ni \psi \mapsto \|T_h \psi - S^\delta\|^2 + \alpha \|\psi\|^2.$$

*If  $\alpha > \eta_h^2$ , then*

$$\beta \|\psi_{\alpha,h}^\delta - \psi^\dagger\|^2 \leq \frac{6\eta_h^2 \|\psi^\dagger\|^2 + 4\delta^2}{\alpha - \eta_h^2} + (-\phi)^* \left( -\frac{1}{4(\alpha - \eta_h^2)} \right),$$

*where  $(-\phi)^*(s) := \sup_{t \in \mathbb{R}_+} (st + \phi(t))$  denotes the Fenchel conjugate of  $-\phi$ .*

*Proof.* Using the triangle inequality, we have

$$\begin{aligned}
 & \frac{1}{4} \|T(\psi_{\alpha,h}^\delta - \psi^\dagger)\|^2 \\
 & \leq \|(T - T_h)\psi_{\alpha,h}^\delta\|^2 + \|T_h\psi_{\alpha,h}^\delta - S^\delta\|^2 + \|T_h\psi^\dagger - S^\delta\|^2 + \|(T - T_h)\psi^\dagger\|^2 \\
 & \leq \eta_h^2(\|\psi_{\alpha,h}^\delta\|^2 + \|\psi^\dagger\|^2) + \|T_h\psi_{\alpha,h}^\delta - S^\delta\|^2 + \|T_h\psi^\dagger - S^\delta\|^2.
 \end{aligned}$$

Due to this and the minimization property of  $\psi_{\alpha,h}^\delta$ ,

$$\begin{aligned}
 & \alpha\|\psi\|^2 - \alpha\|\psi^\dagger\|^2 \leq \|T_h\psi^\dagger - S^\delta\|^2 - \|T_h\psi_{\alpha,h}^\delta - S^\delta\|^2 \\
 & \leq 2\|T_h\psi^\dagger - S^\delta\|^2 + \eta_h(\|\psi^\dagger\|^2 + \|\psi_{\alpha,h}^\delta\|^2) - \frac{1}{4}\|T(\psi_{\alpha,h}^\delta - \psi^\dagger)\|^2 \\
 & \leq 2(\eta_h\|\psi^\dagger\| + \delta)^2 + \eta_h^2(\|\psi^\dagger\|^2 + \|\psi_{\alpha,h}^\delta\|^2) - \frac{1}{4}\|T(\psi_{\alpha,h}^\delta - \psi^\dagger)\|^2 \\
 & \leq 5\eta_h^2\|\psi^\dagger\|^2 + 4\delta^2 + \eta_h^2\|\psi_{\alpha,h}^\delta\|^2 - \frac{1}{4}\|T(\psi_{\alpha,h}^\delta - \psi^\dagger)\|^2.
 \end{aligned}$$

Let  $\gamma > 0$  be arbitrary. Inserting the above inequality into assumption (9) yields

$$\begin{aligned}
 & \beta\|\psi_{\alpha,h}^\delta - \psi^\dagger\| \\
 & \leq \gamma(\|\psi^\dagger\|^2 - \|\psi_{\alpha,h}^\delta\|^2) + (1 + \gamma)(\|\psi_{\alpha,h}^\delta\|^2 - \|\psi^\dagger\|^2) + \phi(\|T(\psi_{\alpha,h}^\delta - \psi^\dagger)\|^2) \\
 & \leq \gamma\|\psi^\dagger\|^2 + \frac{1 + \gamma}{\alpha}(5\eta_h^2\|\psi^\dagger\|^2 + 4\delta^2) + \left(\frac{(1 + \gamma)\eta_h^2}{\alpha} - \gamma\right)\|\psi_{\alpha,h}^\delta\|^2 \\
 & \quad - \frac{1 + \gamma}{4\alpha}\|T(\psi_{\alpha,h}^\delta - \psi^\dagger)\|^2 + \phi(\|T(\psi_{\alpha,h}^\delta - \psi^\dagger)\|^2) \\
 & \leq \gamma\|\psi^\dagger\|^2 + \frac{(1 + \gamma)}{\alpha}(5\eta_h^2\|\psi^\dagger\|^2 + 4\delta^2) + \left(\frac{(1 + \gamma)\eta_h^2}{\alpha} - \gamma\right)\|\psi_{\alpha,h}^\delta\|^2 \\
 & \quad + (-\phi)^*\left(-\frac{1 + \gamma}{4\alpha}\right)
 \end{aligned}$$

If we choose  $\gamma$  such that

$$\frac{(1 + \gamma)\eta_h^2}{\alpha} - \gamma = 0,$$

which requires  $\alpha > \eta_h^2$  in order for  $\gamma$  to be positive, we finally obtain the desired estimate.  $\square$

Using an a-priori parameter choice similar to [4], which for differentiable  $\phi$  amounts to

$$\alpha(\delta, h) \sim \eta_h^2 + \frac{1}{\phi'(\eta_h^2\|\psi^\dagger\|^2 + \delta^2)} > \eta_h^2,$$

leads to a convergence rate

$$\|\psi_{\alpha,h}^\delta - \psi^\dagger\|^2 = \mathcal{O}(\phi(\eta_h^2\|\psi^\dagger\|^2 + \delta^2)).$$

For  $C = X$  and  $\phi(t) \sim \sqrt{t}$ , which corresponds to  $\mu = \frac{1}{2}$  in (7), this rate is identical to (8) with respect to the data error  $\delta$ , but somewhat worse with respect to the operator error  $\eta_h$ .

## 4 Properties of the space $H_{\text{fib}}(\Omega \times S^2)$

What remains to analyze is the convergence  $T_h \rightarrow T$ . As mentioned before, this would not be possible if  $T$  was defined on  $L^2(\Omega \times S^2)$ . In the following, we will show that  $H_{\text{fib}}(\Omega \times S^2)$  is compactly embedded in  $L^2(\Omega \times S^2)$ , thus enabling convergence if the projections converge pointwise.

We introduce the operator  $D_0: H_{\text{fib}}(\Omega \times S^2) \subset L^2(\Omega \times S^2) \rightarrow L^2(\Omega \times S^2)^3 \times L^2(\Omega \times S^2)^3$  by

$$D_0 := \begin{pmatrix} D_{\text{fib}} \\ \text{Grad} \end{pmatrix}$$

in order to write

$$\|\psi\|_{\text{fib}}^2 = \|\psi\|^2 + \|D_0\psi\|^2.$$

Moreover, we will need Sobolev spaces on  $\Omega \times S^2$ . For this purpose, define the operator  $\Lambda^s$  for  $s \in \mathbb{R}$  by

$$\Lambda^s \psi := \sum_{klm} \lambda_{kl}^s \langle \psi, \eta_{klm} \rangle \eta_{klm}, \quad \lambda_{kl} = \sqrt{1 + \|k\|^2 + l^2},$$

whenever the series converges. Then the Sobolev space  $H^s(\Omega \times S^2)$  of order  $s \geq 0$  can be written as the set of all  $\psi \in L^2(\Omega \times S^2)$  for which  $\Lambda^s \psi \in L^2(\Omega \times S^2)$ , equipped with the inner product

$$\langle \psi, \phi \rangle_{H^s(\Omega \times S^2)} = \langle \Lambda^s \psi, \Lambda^s \phi \rangle.$$

It is easy to see that a function  $\psi \in L^2(\Omega \times S^2)$  belongs to  $H^1(\Omega \times S^2)$  if and only if both  $\text{grad } \psi$  and  $\text{Grad } \psi$  exist in  $L^2(\Omega \times S^2)$ , and that there is an equivalence of norms

$$\|\cdot\|_{H^1(\Omega \times S^2)}^2 \simeq \|\cdot\|^2 + \|\text{grad } \cdot\|^2 + \|\text{Grad } \cdot\|^2.$$

We first need some auxiliary results.

**Lemma 4.1.**  *$H_{\text{fib}}(\Omega \times S^2)$  equipped with the norm  $\|\cdot\|_{\text{fib}}$  is a Hilbert space. It coincides with the domain of definition of the self-adjoint operator  $|D_0| := (D_0^* D_0)^{1/2}$ .*

*Proof.*  $D_{\text{fib}}$  is closed, as follows from the weak characterization

$$D_{\text{fib}} \psi = \phi \iff \langle \phi, g \rangle = -\langle \psi, D_{\text{fib}} g \rangle \quad \text{for all } g \in C_{\text{per}}^\infty(\Omega \times S^2).$$

This and the closedness of  $\text{Grad}$  imply that  $D_0$  is closed as well. Therefore, its domain  $\mathcal{D}(D_0) = H_{\text{fib}}(\Omega \times S^2)$  is a Hilbert space with the graph norm  $\|\cdot\|_{\text{fib}} = (\|\cdot\|^2 + \|D_0 \cdot\|^2)^{\frac{1}{2}}$ .

Self-adjointness of  $D_0^* D_0$  (initially defined for example on  $C_{\text{per}}^\infty(\Omega \times S^2)$ ) can be established by a Friedrich's extension (see e.g. [16, Appendix A.8]). Hence  $|D_0|$  is well defined by the functional calculus. By the polar decomposition of closed operators in Hilbert spaces we have  $\mathcal{D}(|D_0|) = \mathcal{D}(D_0)$ .  $\square$

**Lemma 4.2.**  *$C_{\text{per}}^\infty(\Omega \times S^2)$  and  $\text{span}\{\eta_{klm}\}$  are dense in  $H_{\text{fib}}(\Omega \times S^2)$ .*

*Proof.* Since  $\text{span}\{\eta_{klm}\} \subset C_{\text{per}}^\infty(\Omega \times S^2)$ , it suffices to consider  $\text{span}\{\eta_{klm}\}$ .

In a first step, we show that  $\text{span}\{\eta_{klm}\}$  is dense in  $\mathcal{D}(|D_0|^2) \subset H_{\text{fib}}(\Omega \times S^2)$ . For  $n \in \mathbb{N}$ , let the orthogonal projections  $\tau_n: L^2(\Omega \times S^2) \rightarrow L^2(\Omega \times S^2)$  be given by

$$\tau_n \psi := \sum_{klm}^n \langle \psi, \eta_{klm} \rangle_{L^2} \eta_{klm},$$

where the sum runs over all  $(k, l, m)$  with  $|k| \leq n$ ,  $|l| \leq n$  and  $|m| \leq l$ . Then  $\tau_n \psi \rightarrow \psi$  in  $L^2(\Omega \times S^2)$  as  $n \rightarrow \infty$  since the orthonormal system  $\{\eta_{klm}\}$  is complete. It suffices to show that

$$\lim_{n \rightarrow \infty} \|\tau_n \psi - \psi\|_{\text{fib}} = 0 \quad \text{for all } \psi \in \mathcal{D}(|D_0|^2) \subset H_{\text{fib}}(\Omega \times S^2). \quad (10)$$

For all  $\phi \in C_{\text{per}}^\infty(\Omega \times S^2)$ ,

$$\langle |D_0| \tau_n \psi, \phi \rangle \rightarrow \langle |D_0| \psi, \phi \rangle,$$

and since  $C_{\text{per}}^\infty(\Omega \times S^2)$  is dense in  $L^2(\Omega \times S^2)$ , it follows that  $|D_0| \tau_n \psi \rightharpoonup |D_0| \psi$ . Moreover,  $|D_0|^2 \tau_n \phi \rightarrow |D_0|^2 \phi$ . This can be seen using  $|D_0|^2 = -\Delta_{S^2} - D_{\text{fib}}^2$ ,

$$\Delta_{S^2} \tau_n \phi = \tau_n \Delta_{S^2} \phi \rightarrow \Delta_{S^2} \phi$$

and

$$D_{\text{fib}}^2 \tau_n \phi = \sum_{i,j=1}^3 u_i u_j (\tau_n \partial_{x_i} \partial_{x_j} \phi) \rightarrow \sum_{i,j=1}^3 u_i u_j \partial_{x_i} \partial_{x_j} \phi = D_{\text{fib}}^2 \phi,$$

where we have used that both spatial and angular derivatives commute with  $\tau_n$ . Therefore,

$$\langle \tau_n |D_0|^2 \tau_n \psi, \phi \rangle = \langle \tau_n \psi, |D_0|^2 \tau_n \phi \rangle \rightarrow \langle \psi, |D_0|^2 \phi \rangle = \langle |D_0|^2 \psi, \phi \rangle,$$

i.e.  $\tau_n |D_0|^2 \tau_n \psi \rightharpoonup |D_0|^2 \psi$ , and hence  $\||D_0| \tau_n \psi\| \rightarrow \||D_0| \psi\|$ . Since weak convergence and convergence of the norms implies strong convergence, we obtain (10).

The general case can be reduced to the special case above by density of  $\mathcal{D}(|D_0|^2)$  in  $H_{\text{fib}}(\Omega \times S^2) = \mathcal{D}(|D_0|)$  shown below: For  $\psi \in \mathcal{D}(|D_0|)$  and  $\epsilon > 0$  there exists  $\psi_\epsilon \in \mathcal{D}(|D_0|^2)$  such that  $\|\psi - \psi_\epsilon\|_{\text{fib}} \leq \epsilon/2$  and  $n \in \mathbb{N}$  such that  $\|\psi_\epsilon - \tau_n \psi_\epsilon\| \leq \epsilon/2$ . Hence, by the triangle inequality  $\|\psi - \tau_n \psi_\epsilon\| \leq \epsilon$ .

The density of  $\mathcal{D}(|D_0|^2)$  in  $H_{\text{fib}}(\Omega \times S^2) = \mathcal{D}(|D_0|)$  can be shown by considering the spectral decomposition  $(E_\lambda)_{\lambda \in \mathbb{R}}$  of  $|D_0|$ . For  $\psi \in \mathcal{D}(|D_0|)$ , let

$$\psi_n = E_n \psi - E_{-n} \psi, \quad n \in \mathbb{N}.$$

Then

$$\||D_0|^2 \psi_n\| = \int_{-n-}^{n+} \lambda^4 d\|E_\lambda \psi\|^2 \leq n^4 \|\psi\|^2,$$

i.e.  $\psi_n \in \mathcal{D}(|D_0|^2)$ . Moreover,  $\psi_n \rightarrow \psi$  and  $|D_0| \psi_n = E_n |D_0| \psi - E_{-n} |D_0| \psi \rightarrow |D_0| \psi$ .  $\square$

Additionally, some regularity properties of the operators introduced above and their commutators are needed. These could be obtained easily using the formalism of pseudo-differential operators. However, for the special cases considered here, they can also be proved directly.

**Lemma 4.3.** *For each  $s \in \mathbb{R}$ , the operators*

- $\text{grad}, \text{Grad}, D_{\text{fib}}: H^s(\Omega \times S^2) \rightarrow H^{s-1}(\Omega \times S^2)$  and their adjoints,
- $[\Lambda^{-1}, u_i]: H^s(\Omega \times S^2) \rightarrow H^{s+2}(\Omega \times S^2)$ ,
- and  $[\Lambda^{-1}, \text{grad}], [\Lambda^{-1}, \text{Grad}_i]: H^s(\Omega \times S^2) \rightarrow H^{s+1}(\Omega \times S^2)$

are bounded. Here,  $[A, B] := AB - BA$  denotes the commutator.



*Proof.* The first statement is obvious from the definitions. For the second one,

$$[\Lambda^{-1}, u_i]\psi = \sum_k \sum_{lm} \sum_{l'm'} (\lambda_{kl}^{-1} - \lambda_{kl'}^{-1}) \langle \psi, \eta_{kl'm'} \rangle \langle u_i Y_{l'm'}, Y_{lm} \rangle \eta_{klm}.$$

Since  $u_i$  can be written as a linear combination of spherical harmonics of degree 1,  $\langle u_i Y_{lm}, Y_{l'm'} \rangle$  vanishes if  $|l - l'| > 1$ . Using this, we obtain that

$$|\lambda_{kl}^{-1} - \lambda_{kl'}^{-1}| = \mathcal{O}(l \lambda_{kl}^{-3}) \quad \text{as } k, l \rightarrow \infty.$$

Together with the boundedness of  $\langle u_i Y_{lm}, Y_{l'm'} \rangle$ , the assertion follows. Similarly,

$$[\Lambda^{-1}, \text{Grad}_i]\psi = \sum_k \sum_{lm} \sum_{l'm'} (\lambda_{kl}^{-1} - \lambda_{kl'}^{-1}) \langle \psi, \eta_{kl'm'} \rangle \langle \text{Grad}_i Y_{l'm'}, Y_{lm} \rangle \eta_{klm}.$$

Again,  $\langle \text{Grad}_i Y_{l'm'}, Y_{lm} \rangle$  vanishes for  $|l - l'| > 1$  and behaves like  $\mathcal{O}(l)$  for  $l \rightarrow \infty$ , as can be seen from

$$\begin{aligned} \sum_i |\langle \text{Grad}_i Y_{l'm'}, Y_{lm} \rangle|^2 &\leq \sum_i \|\text{Grad}_i Y_{l'm'}\|^2 \|Y_{lm}\|^2 = \sum_i \|\text{Grad}_i Y_{l'm'}\|^2 \\ &= \|\text{Grad } Y_{l'm'}\|^2 = l'(l' + 1). \end{aligned}$$

Together with  $[\Lambda^{-1}, \text{grad}] = 0$ , this shows the last assertion.  $\square$

We can now prove the embedding theorem. The proof uses techniques related to the proof of Hörmanders theorem on hypo-elliptic operators (c.f. [11]) found in [9].

**Theorem 4.4.**  $H_{\text{fib}}(\Omega \times S^2)$  is a continuously embedded subspace of  $H^{1/2}(\Omega \times S^2)$ .

*Proof.* Assume first that  $\psi \in C_{\text{per}}^\infty(\Omega \times S^2)$ . Let  $D_1 := [\text{Grad}, D_{\text{fib}}]$ . Then, since  $\text{Grad}(a^T \cdot)(u) = (1 - uu^T)a$  for  $a \in \mathbb{R}^3$ , we have

$$\begin{aligned} D_1 \psi &= \sum_{i=1}^3 (\text{Grad } u_i) \text{grad}_i \psi \\ &= (1 - uu^T) \text{grad } \psi. \end{aligned}$$

This implies

$$\|D_{\text{fib}} \psi\|^2 + \|D_1 \psi\|^2 = \|\text{grad } \psi\|^2$$

and hence

$$\Lambda^2 = 1 + D_0^* D_0 + D_1^* D_1.$$

Thus

$$\begin{aligned} \|\Lambda^{1/2} \psi\|^2 &= \langle \psi, \Lambda^{-1} \Lambda^2 \psi \rangle = \|\Lambda^{-1/2} \psi\|^2 + \sum_{i=0}^1 \langle \psi, \Lambda^{-1} D_i^* D_i \psi \rangle \\ &= \|\Lambda^{-1/2} \psi\|^2 + \sum_{i=0}^1 \langle D_i \psi, \Lambda^{-1} D_i \psi \rangle + \langle \psi, [\Lambda^{-1}, D_i^*] \Lambda^{1/2} \Lambda^{-1/2} D_i \psi \rangle \\ &\leq \|\psi\|^2 + \sum_{i=0}^1 \|\Lambda^{-1/2} D_i \psi\|^2 + c \|\psi\| \|\Lambda^{-1/2} D_i \psi\| \\ &\leq c (\|\psi\| + \|D_0 \psi\| + \|\Lambda^{-1/2} D_1 \psi\|)^2, \end{aligned} \tag{11}$$

where all generic constants are denoted by  $c$  for simplicity. We used that  $\Lambda^{-1/2}$  and  $[\Lambda^{-1}, D_i^*]\Lambda^{1/2}$  ( $i = 1, 2$ ) are bounded. The latter follows using the regularity properties in Lemma 4.3, the Leibniz rule for commutators, i.e.  $[A, BC] = [A, B]C + B[A, C]$ , and the expression

$$\text{Grad}_i^* \psi = -\text{Grad}_i \psi + 2u_i \psi, \quad (12)$$

for the (formal) adjoint of the  $i$ -th component of  $\text{Grad}$ . This in turn can be deduced using partial integration and the relation

$$\begin{aligned} \int_{S^2} \text{Grad}_i f(u) du &= \int_{S^2} (\text{Grad } u_i)^T \text{Grad } f(u) du \\ &= - \int_{S^2} (\Delta_{S^2} u_i) f(u) du = 2 \int_{S^2} u_i f(u) du, \end{aligned}$$

where  $\Delta_{S^2} = -\text{Grad}^* \circ \text{Grad}$  and  $\Delta_{S^2} u_i = -2u_i$  was used. For the  $D_1$ -term in (11), we have

$$\begin{aligned} \|\Lambda^{-1/2} D_1 \psi\|^2 &= \sum_{i=1}^3 \langle \Lambda^{-1} D_{1,i} \psi, D_{1,i} \psi \rangle \\ &\leq \sum_{i=1}^3 |\langle \Lambda^{-1} D_{1,i} \psi, D_{\text{fib}} \text{Grad}_i \psi \rangle| + |\langle \Lambda^{-1} D_{1,i} \psi, \text{Grad}_i D_{\text{fib}} \psi \rangle|. \end{aligned}$$

The first term of this can be estimated by

$$\begin{aligned} &|\langle \Lambda^{-1} D_{1,i} \psi, D_{\text{fib}} \text{Grad}_i \psi \rangle| \\ &\leq |\langle \Lambda^{-1} D_{1,i} D_{\text{fib}}^* \psi, \text{Grad}_i \psi \rangle| + |\langle [D_{\text{fib}}^*, \Lambda^{-1} D_{1,i}] \psi, \text{Grad}_i \psi \rangle| \\ &\leq c(\|D_{\text{fib}}^* \psi\| \|\text{Grad}_i \psi\| + \|\psi\| \|\text{Grad}_i \psi\|) \\ &\leq c(\|\psi\| + \|D_0 \psi\|)^2, \end{aligned}$$

where where  $D_{\text{fib}}^* = -D_{\text{fib}}$  and the boundedness of  $\Lambda^{-1} D_{1,i}$  and  $[D_{\text{fib}}^*, \Lambda^{-1} D_{1,i}]$  was used. Similarly,

$$\begin{aligned} &|\langle \Lambda^{-1} D_{1,i} \psi, \text{Grad}_i D_{\text{fib}} \psi \rangle| \\ &\leq |\langle \Lambda^{-1} D_{1,i} \text{Grad}_i^* \psi, D_{\text{fib}} \psi \rangle| + |\langle [\text{Grad}_i^*, \Lambda^{-1} D_{1,i}] \psi, D_{\text{fib}} \psi \rangle| \\ &\leq c(\|\text{Grad}_i^* \psi\| \|D_{\text{fib}} \psi\| + \|\psi\| \|D_{\text{fib}} \psi\|), \end{aligned}$$

since  $[\text{Grad}_i^*, \Lambda^{-1} D_{1,i}]$  is bounded. Using (12) again, the commutator between  $\text{Grad}_i$  and its adjoint is

$$[\text{Grad}_i, \text{Grad}_i^*] \psi = 2(1 - u_i^2) \psi.$$

Hence,

$$|\|\text{Grad}_i^* \psi\|^2 - \|\text{Grad}_i \psi\|^2| = |\langle \psi, [\text{Grad}_i^*, \text{Grad}_i] \psi \rangle| \leq 2\|\psi\|^2,$$

so the second term can be estimated further by

$$|\langle \Lambda^{-1} D_{1,i} \psi, \text{Grad}_i D_{\text{fib}} \psi \rangle| \leq c(\|\psi\| + \|D_0 \psi\|)^2.$$

Putting everything together, we obtain

$$\|\psi\|_{H^{1/2}(\Omega \times S^2)} = \|\Lambda^{1/2} \psi\| \leq c\|\psi\|_{\text{fib}}$$

for  $\psi \in C_{\text{per}}^\infty(\Omega \times S^2)$ . Finally, Lemma 4.2 implies that the estimate also holds for arbitrary  $\psi \in H_{\text{fib}}(\Omega \times S^2)$ .  $\square$

**Corollary 4.5.**  $H_{\text{fib}}(\Omega \times S^2)$  is compactly embedded in  $L^2(\Omega \times S^2)$ .

*Proof.* This follows immediately from the fact that  $H^s(\Omega \times S^2)$  is compactly embedded in  $L^2(\Omega \times S^2)$  for any  $s > 0$ .  $\square$

In particular,  $T$  is compact since it is continuous on  $L^2(\Omega \times S^2)$ . Hence, the following holds true:

**Corollary 4.6.**  $\|T - T_h\| \rightarrow 0$  if and only if  $P_h \rightarrow 1$  pointwise on  $\mathcal{R}(T^*)$  and  $Q_h \rightarrow 1$  pointwise on  $\mathcal{R}(T)$ .

Hence, under the assumptions of Theorem 3.1 and Corollary 4.6 and a parameter choice rule  $\alpha(\delta, h)$  satisfying  $\alpha(\delta, h) \geq \|T - T_h\|^2$ , it follows that

$$\|\psi_{\alpha(\delta, h), h}^\delta - \psi^\dagger\|_{\text{fib}} \rightarrow 0 \quad \text{as } \delta, h \rightarrow 0.$$

## 5 Implementation

For the numerical implementation, we use the basis introduced in (4) above<sup>1</sup> and discretize by truncating to  $(\eta_{klm})$  for  $|k| \leq K_h$  and  $l \leq L_h$ . The same basis is used in data space. Let  $P_h$  and  $Q_h$  denote the orthogonal projections onto  $X_h := \text{span}\{\eta_{klm} : |k| \leq K_h, l \leq L_h\}$  with respect to the norms  $\|\cdot\|_{\text{fib}}$  and  $\|\cdot\|_{L^2}$ , respectively. The pointwise convergence  $\|P_h\psi - \psi\|_{\text{fib}} \rightarrow 0$  for all  $\psi \in H_{\text{fib}}(\Omega \times S^2)$  and  $\|Q_h\psi - \psi\|_{\text{fib}} \rightarrow 0$  for all  $\psi \in L^2(\Omega \times S^2)$  follows from Lemma 4.2 if  $K_h, L_h \rightarrow \infty$  as  $h \rightarrow 0$ . We will need the Gramian matrix

$$G_{klm, k'l'm'} = \langle \eta_{klm}, \eta_{k'l'm'} \rangle_{\text{fib}}$$

for this basis. The computation of  $G$  requires evaluation of integrals of the form

$$\int_{S^2} u_i u_j Y_{lm}(u) \overline{Y_{l'm'}(u)} du.$$

These can be calculated either by numerical quadrature, or using explicit — albeit rather tedious — formulas for Clebsch-Gordan coefficients.

We first numerically test the convergence of  $P_h$ . Let  $\iota: H_{\text{fib}}(\Omega \times S^2) \rightarrow L^2(\Omega \times S^2)$  be the embedding. Then we want to investigate the norm convergence  $\|\iota \circ P_h - \iota\| \rightarrow 0$  as  $h \rightarrow 0$ . The problem here is that for  $\psi \in H_{\text{fib}}(\Omega \times S^2)$ , one needs to quantify  $\psi - P_h\psi$ , i.e. its component outside of the discrete subspace. We approximate this by choosing a fine discretization  $X_0 := \text{span}\{\eta_{klm} : |k| \leq K_0, l \leq L_0\}$  for some sufficiently large  $K_0, L_0$  and use this in place of the infinite-dimensional  $H_{\text{fib}}(\Omega \times S^2)$ . Then  $P_h$  and the norm above can be evaluated numerically by taking advantage of the fact that the Gramian matrices of  $X_0$  and  $X_h$  decouple with respect to the spatial part  $\xi_k$  of the basis functions. In Figure 1, the results for various  $L_h$  are plotted against the spatial frequency cutoff  $K_h$  for norms with and without the additional Laplace-Beltrami operator, clearly showing convergence almost independently of the chosen  $L_h$ . The figure also shows the results of the same numerical test when omitting the Laplace-Beltrami penalty. While these also decrease for large enough  $K_h$ , the effect becomes weaker as the SH order  $L_h$  increases. This suggests that the observed decrease can be seen as an artifact

<sup>1</sup>More precisely, we use a slightly modified basis consisting of real-valued linear combinations of (4).

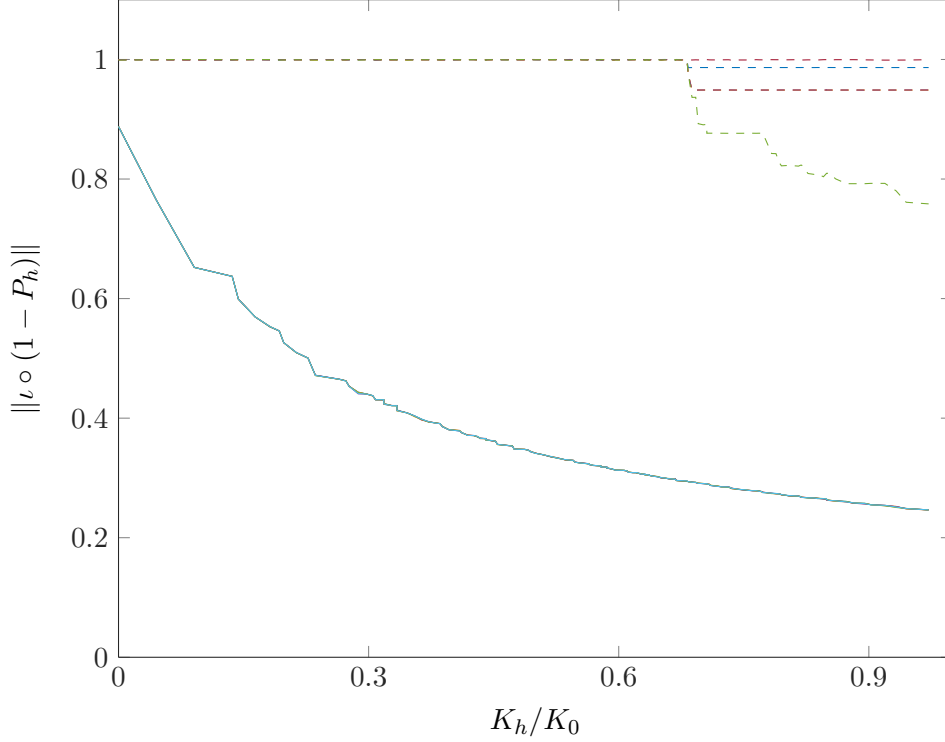


Figure 1: Convergence of the projection  $P_h$  viewed as a map  $H_{\text{fib}}(\Omega \times S^2) \rightarrow L^2(\Omega \times S^2)$  ( $\iota$  is the embedding) plotted against the spatial cutoff frequency for even SH orders  $10 \leq L_h \leq 20$ . *Solid*: Including the angular regularization term in (5) (the curves are almost indistinguishable). *Dashed*: Without the angular regularization term; SH order is increasing from bottom up.

of the Spherical Harmonics themselves introducing a smoothing effect (i.e. regularization by discretization), and hence convergence speed with respect to  $K_h$  deteriorates with increasing  $L_h$ . Moreover, this might explain why good results can be observed even without additional explicit angular regularization.

For ODF reconstruction, the projected ODF and the forward operator are expanded in the discrete basis as

$$P_h \psi = \sum_{klm} z_{klm} \eta_{klm}$$

and

$$T_h \eta_{klm} = Q_h T \eta_{klm} = \sum_{k'l'm'} (B_h)_{k'l'm',klm} \eta_{k'l'm'}.$$

For the spherical convolution model,  $B_h$  is diagonal. Similarly, the discrete data is expanded as

$$Q_h S^\delta = \sum_{klm} y_{klm}^\delta \eta_{klm}.$$

Note that this can only be approximated in practice.

A method to implement the constrained spherical deconvolution problem (6) was proposed in [18]. In the following, we will describe this method and interpret it as a semi-smooth Newton

method. This will prove local convergence of the method by general convergence results for semi-smooth Newton methods.

Instead of requiring  $P_h\psi \geq 0$ , the constraint is checked only on a finite subset  $\{(x_i, u_j)\}_{ij} \subset \Omega \times S^2$ ,

$$(Hz)_{ij} := \sum_{klm} z_{klm} \eta_{klm}(x_i, u_j) \geq 0.$$

The points  $x_i$  are usually chosen as the spatial grid on which the data  $S^\delta$  are given, while  $\{u_j\}$  is some point set on the sphere obtained e.g. by subdividing platonic solids. The constraint is only implemented approximately by an iteration of the form

$$z_{k+1} = \underset{z}{\operatorname{argmin}} \left( \|B_h z - y^\delta\|^2 + \alpha \|\sqrt{G}z\|^2 + c \|\theta(-Hz_k) \cdot Hz\|^2 \right), \quad (13)$$

where  $\theta$  is the Heaviside step function and  $\cdot$  denotes element-wise multiplication. The constraint is approximated by a quadratic penalty on the set where the previous iterate violated it, i.e.  $Hz_k < 0$ . The parameter  $c > 0$  determines how strongly the constraint is enforced. This method is equivalent to applying the semi-smooth Newton method from [8] to the relaxed problem

$$\underset{Hz \geq w}{\operatorname{argmin}} \left( \|B_h z - y^\delta\|^2 + \alpha \|\sqrt{G}z\|^2 + c \|w\|^2 \right). \quad (14)$$

The equivalence can be seen by noting that the necessary and sufficient first order optimality conditions for (14) can be written as (see the reference cited above for more details)

$$\begin{aligned} (B_h^* B_h + \alpha G)z - H^* \lambda &= B_h^* y^\delta \\ cw + \lambda &= 0 \\ \lambda &= \max(0, \lambda - c(Hz - w)), \end{aligned}$$

which is the same as

$$(B_h^* B_h + G)z + cH^* \min(0, Hz) = B_h^* y^\delta.$$

Finally, (13) is precisely a semi-smooth Newton step for this equation. The minimization problem (14) can be interpreted as Moreau-Yosida regularization of the constraint  $Hz \geq 0$ . Convergence of both the semi-smooth Newton method for  $k \rightarrow \infty$  and the Moreau-Yosida regularization for  $c \rightarrow \infty$  have been investigated in [8, 12]. For the numerical experiments, we will use this method since it is efficient and easy to implement. It allows for comparison of the regularized method with the original CSD method, which is included as the case  $\alpha = 0$ .

## 6 Numerical experiments

Performance of the method was tested on a physical phantom and an *in vivo* measurement. Both data sets were obtained by other groups.

Construction of the phantom and data acquisition is described in more detail in [7]. The data was acquired on a  $64 \times 64 \times 3$ -grid with 3 mm isotropic voxel spacing, using 64 gradient directions and  $b = 1500 \text{ s/mm}^2$ . The structure of the phantom is shown in figure 2, together with three regions of interest — two crossings and a curved structure. Reconstructions were performed on the whole data set, without employing a “white” matter mask. The SH coefficients of the convolution kernel up to order 8 were estimated from the data using



Figure 2: Structure of the physical phantom ( $b = 0$  image). Reconstructions for the highlighted regions are depicted in figure 3.

the *MRtrix* software package.<sup>2</sup> The ODF was reconstructed in Spherical Harmonics up to order 12. The parameter  $c$  was simply chosen as  $c = 1$ . Figure 3 shows reconstruction results for these regions using four regularization strategies: unregularized constrained spherical deconvolution (CSD), deconvolution with fiber continuity penalty (CSD-FC), with additional angular Laplace-Beltrami penalty (CSD-FC+LB), and CSD with isotropic spatial regularization (CSD-iso). For the crossing regions, the spatial penalties clearly improve spatial coherence and resolution of the crossing compared to the unregularized reconstruction. Compared to the isotropic penalty, the FC penalty significantly reduces artifacts perpendicular to the fibers. For the curved structure, the FC penalty causes some visible artifacts tangential to the structure. These are reduced somewhat by the additional angular penalty, essentially by blurring the spurious crossings. This of course also limits the achievable angular resolution. It also produces some artifacts perpendicular to the fibers since the embedding into  $H^{1/2}(\Omega \times S^2)$  implies some degree of isotropic smoothness. Therefore, the corresponding regularization parameter should not be chosen too large.

The *in vivo* data set was taken from the Human Connectome Project (HCP) database.<sup>3</sup> The data set consists of each 90 diffusion weighted images for  $b$ -values 1000, 2000 and 3000 s/mm<sup>2</sup>. Of these, we only used the  $b = 2000$  s/mm<sup>2</sup> points. As the data set is intended primarily for brain research, not for evaluating reconstruction and tracking methods, its SNR is rather high, so that even unregularized reconstructions show a good spatial coherence. Still, it is interesting to test the performance the CSD-FC method on this data set, in particular to see how well the geometric assumption of locally straight fibers is fulfilled in a realistic situation. Therefore, the spatial regularization parameter was deliberately chosen rather large to highlight some of its strengths and shortcomings.

Reconstructions were performed on a  $30 \times 30 \times 35$  voxel subset of the volume around the area shown in figure 4. We parametrized the convolution kernel as an exponential function,  $k(t) \simeq \exp(-\alpha t^2)$ , with parameters obtained from a previous DTI reconstruction by averaging the

<sup>2</sup>Available at <http://www.brain.org.au/software/>. Developed at Brain Research Institute, Melbourne, Australia. Further information can also be found in in [19].

<sup>3</sup>See <https://ida.loni.usc.edu> and the *Acknowledgments* section.

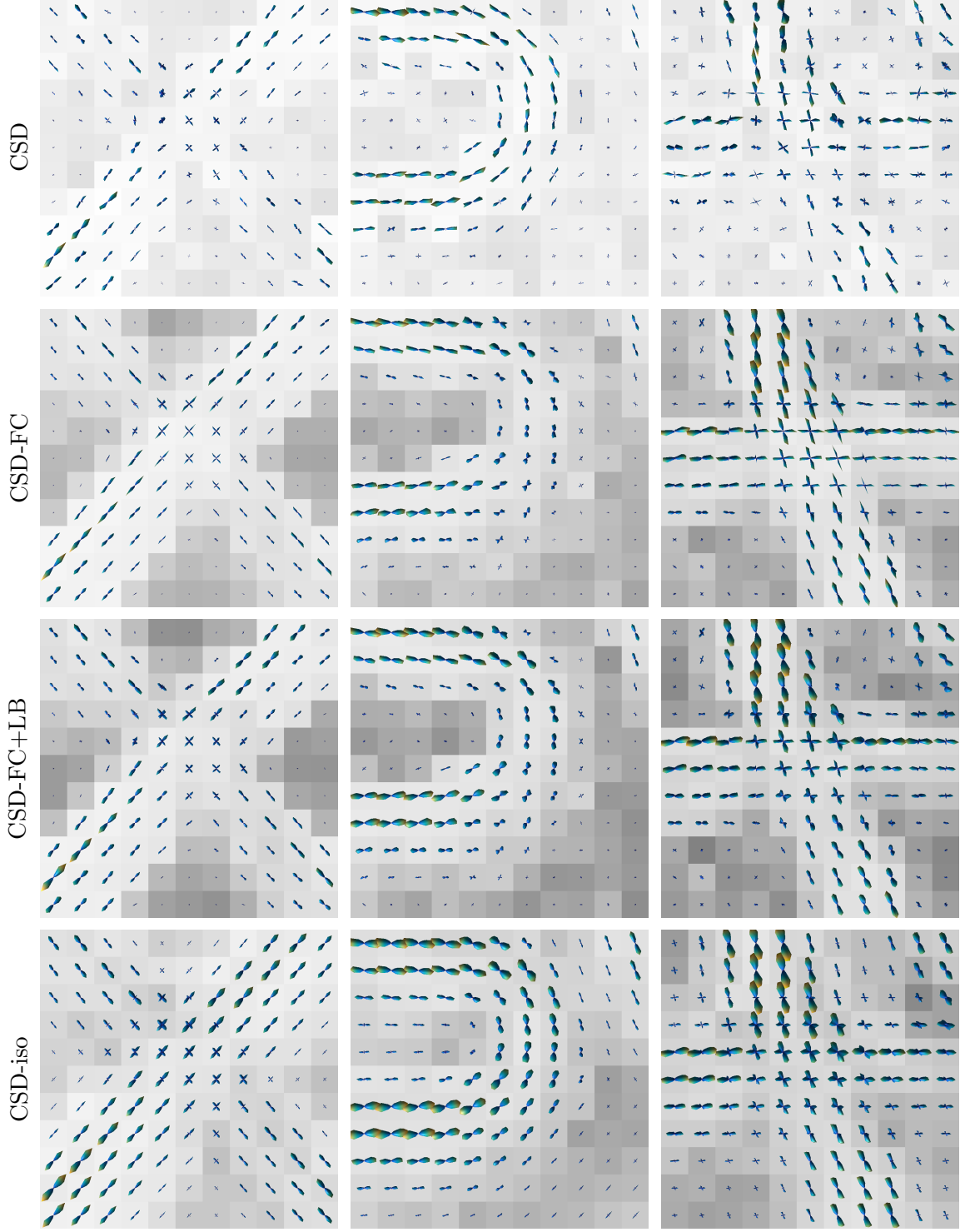


Figure 3: Performance of different regularization strategies on the highlighted regions in figure 2. From top to bottom: (1) Constrained Spherical Deconvolution, (2) CSD with FC penalty (3), CSD with FC and Laplace-Beltrami penalty, (4) CSD with isotropic spatial penalty. The background image shows the (generalized) FA.

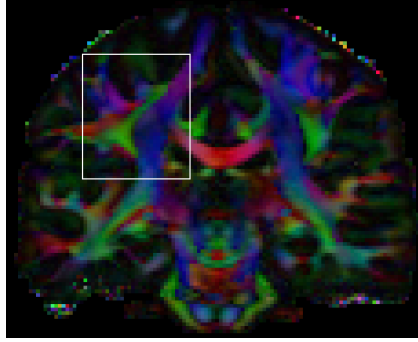


Figure 4: Subregion of the *in vivo* on which reconstructions were performed (the image shows FA color coded by main diffusion direction). The depicted slice is the one in figure 5, while figure 6 is a bit closer to the anterior.

sorted eigenvalues of all tensors with fractional anisotropy (FA) exceeding a certain threshold and taking the mean value of the smaller two eigenvalues in order to obtain an axially symmetric kernel. Instead of an unregularized reconstruction for comparison, we used a small  $L^2$  penalty. Visual inspection showed that this did not have significant impact on the resolved structures. However, it increased FA contrast by reducing noise in empty areas and, more importantly, lead to ODFs that are slightly broader and can therefore be depicted in printed plots more clearly.

Reconstruction results for two slices are shown in figures 5 and 6. For brevity’s sake, we only compare the  $L^2$ -penalized reconstruction to the CSD-FC+LB method.

The top of figure 5 shows a rather complicated “kissing” fiber structure. As in the phantom data set above, the FC penalty tends to introduce spurious crossings in the sharply bent part, which can also be seen in the top highlighted region. The bottom highlighted region counter-intuitively shows a *reduced* quality of the resolved crossing structures. We observed the same effect with spatial regularization alone, without the Laplace-Beltrami penalty. The reason for this observation is not clear.

For the slice in figure 6, the highlighted region at the bottom shows a straight elongated structure, the coherence of which is significantly improved by the FC penalty. The other region shows a white matter structure extending into a gray matter area, which may or may not be an artifact due to the general tendency of the penalty to prolong elongated structures. In cases like these, validation based on other sources of knowledge about the local structures is necessary.

A notable feature of the regularized reconstructions is the significantly improved FA contrast between gray and white matter. In the figures shown, this might be attributed to the Laplace-Beltrami penalty. However, we observed the same effect with spatial regularization alone, i.e. the spatial penalty is able to distinguish between noisy and oriented structures using the fact that the former are not coherent with their surroundings.

## 7 Discussion

In this work, we proved the convergence of fiber continuity based spatial regularization for discrete, noisy data by showing that the natural Sobolev-type space for this regularization



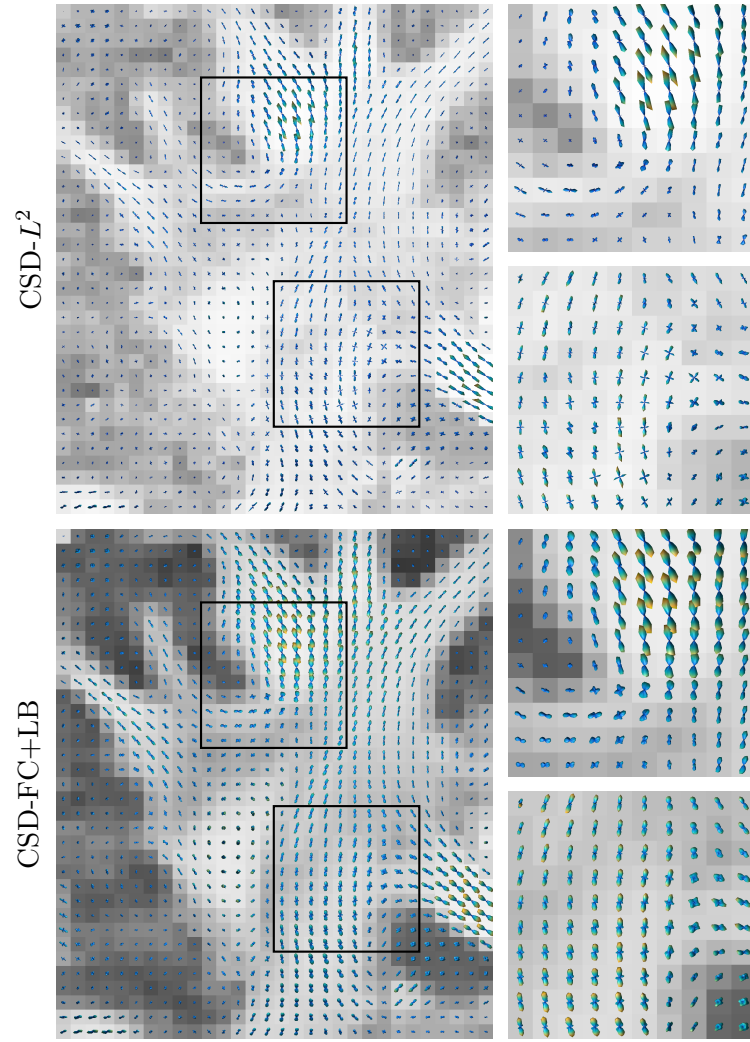


Figure 5: Comparison of CSD with  $L^2$  penalty (top) to CSD with FC and Laplace-Beltrami penalty (bottom) for *in vivo* data shown in Figure 4.

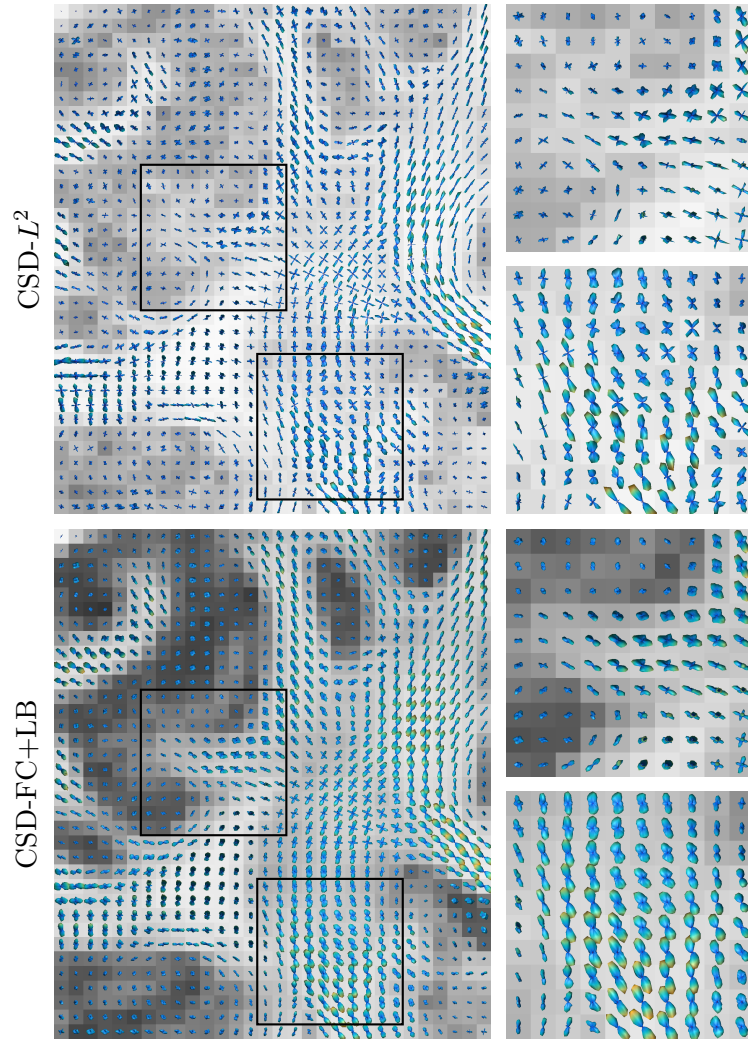


Figure 6: Comparison of CSD with  $L^2$  penalty (top) to CSD with FC and Laplace-Beltrami penalty (bottom) for *in vivo* data shown in Figure 4.

is compactly embedded in  $L^2$ . This allows the approximation of the otherwise non-compact forward operator by finite dimensional operators.

We presented some numerical examples illustrating the performance of the method. They show a significantly more coherent ODF field compared to unregularized reconstructions, the potential to resolve crossing structures even with poor signal to noise ratio, as well as better noise suppression in isotropic areas. It is, however, not generally clear in all cases whether the observed structures are artifacts, e.g. spurious crossings or structures being extended into isotropic regions. Suppression or blurring of existing structures by the method may also occur. In these cases, validation of the results using other methods would be useful. In particular, tractography results from regularized reconstructions would be interesting, as they allow for comparison with existing knowledge about the structure of the brain, in particular when analyzing the widely used HCP data set.

As was already found in [15], the approach shows characteristic weaknesses in curved structures where the smoothness assumption becomes invalid. Both the theoretical results and numerical experiments suggest that the ODF reconstructions can be improved in these areas by inclusion of an additional penalty enforcing angular smoothness. The induced blurring, however, limits angular resolution, which is undesirable for crossings at acute angles. A way out of this may be locally adaptive choices of regularization parameters.

## Acknowledgments

Funding was provided by the DFG Research Training Group 1023 “Identification in Mathematical Models”.

We would like to thank Jens Frahm and Sabine Hofer (Biomedizinische NMR Forschungs GmbH, Göttingen) for helpful discussions.

Data used in the preparation of this work were obtained from the Human Connectome Project (HCP) database. The HCP project (Principal Investigators: Bruce Rosen, M.D., Ph.D., Martinos Center at Massachusetts General Hospital; Arthur W. Toga, Ph.D., University of Southern California, Van J. Weeden, MD, Martinos Center at Massachusetts General Hospital) is supported by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH) and the National Institute of Neurological Disorders and Stroke (NINDS). HCP is the result of efforts of co-investigators from the University of Southern California, Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Washington University, and the University of Minnesota.

## References

- [1] S. BECKER, K. TABELOW, H. U. VOSS, A. ANWANDER, R. M. HEIDEMANN AND J. POLZEHL, *Position-orientation adaptive smoothing of diffusion weighted magnetic resonance data (POAS)*, Med. Imag. Anal., 16 (2012), pp. 1142–1155.
- [2] R. DITS AND E. FRANKEN, *Left-invariant diffusions on the space of positions and orientations and their application to crossing-preserving smoothing of HARDI images*, Int. J. Comput. Vision, 92 (2011), pp. 231–264.
- [3] H. W. ENGL, M. HANKE AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

- [4] J. FLEMMING, *Generalized Tikhonov regularization: basic theory and comprehensive results on convergence rates*, Ph.D. thesis, Chemnitz University of Technology, 2011.
- [5] J. FLEMMING AND B. HOFMANN, *Convergence rates in constrained Tikhonov regularization: equivalence of projected source conditions and variational inequalities*, Inverse Probl., 27 (2011), 085001.
- [6] A. GOH, C. LENGLET, P. M. THOMPSON AND R. VIDAL, *Estimating orientation distribution functions with probability density constraints and spatial regularity*, in Medical Image Computing and Computer-Assisted Intervention — MICCAI 2009, G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble and C. Taylor, eds., Springer-Verlag, 2009, pp. 877–885.
- [7] P. FILLARD, M. DESCOTEAUX *et al.*, *Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom*, Neuroimage, 56 (2011), pp. 220–234
- [8] M. HINTERMÜLLER, K. ITO AND K. KUNISCH, *The primal-dual active set strategy as a semismooth newton method*, SIAM J. Optim., 13 (2003), pp. 865–888
- [9] B. HELFFER AND F. NIER, *Hypoelliptic Estimates and Spectral Theory for Fokker-Planck Operators and Witten Laplacians*, Springer-Verlag, Berlin, 2005.
- [10] B. HOFMANN AND M. YAMAMOTO, *On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems*, Appl. Anal., 89 (2010), pp. 1705–1727.
- [11] L. HÖRMANDER, *Hypoelliptic second order differential equations*, Acta Math., 119 (1967), pp. 147–171
- [12] K. ITO AND K. KUNISCH, *Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces*, Nonlinear Anal., 41 (2000), pp. 591–616
- [13] S. LU AND J. FLEMMING, *Convergence rate analysis of Tikhonov regularization for nonlinear ill-posed problems with noisy operator*, Inverse Probl., 28 (2012), 104003.
- [14] R. PLATO AND G. VAINIKKO, *On the regularization of projection methods for solving ill-posed problems*, Numer. Math., 57 (1990), pp. 63–79.
- [15] M. REISERT AND V. G. KISELEV, *Fiber continuity: An anisotropic prior for ODF estimation*, IEEE Trans. Med. Imaging, 30 (2011), pp. 1274–1283.
- [16] M. TAYLOR *Partial Differential Equations: Basic Theory*, Springer, New York, 1996.
- [17] J. TOURNIER, F. CALAMANTE, D. G. GADIAN AND A. CONNELLY, *Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution*, NeuroImage, 23 (2004), pp. 1176–1185.
- [18] J. TOURNIER, F. CALAMANTE AND A. CONNELLY, *Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution*, NeuroImage, 35 (2007), pp. 1459–1472.
- [19] J. TOURNIER, F. CALAMANTE AND A. CONNELLY, *MRtrix: Diffusion tractography in crossing fiber regions*, Int. J. Imag. Syst. Tech., 22 (2012), pp. 53–66.